Lead Scoring Case Study using Logistic Regression

SUBMITTED BY :

Bariki Pravalika

Contents

- Problem statement
- Problem approach
- EDA
- Correlations
- Model Evaluation
- Observations
- Conclusion

Problem Statement

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have process of form filling on their website after which the company that individual as a lead.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.Through this process, some of the leads get converted while most do not.
- The typical lead conversion rate at X education is around **30%.** Now, this means if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as Hot Leads.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Business Objective

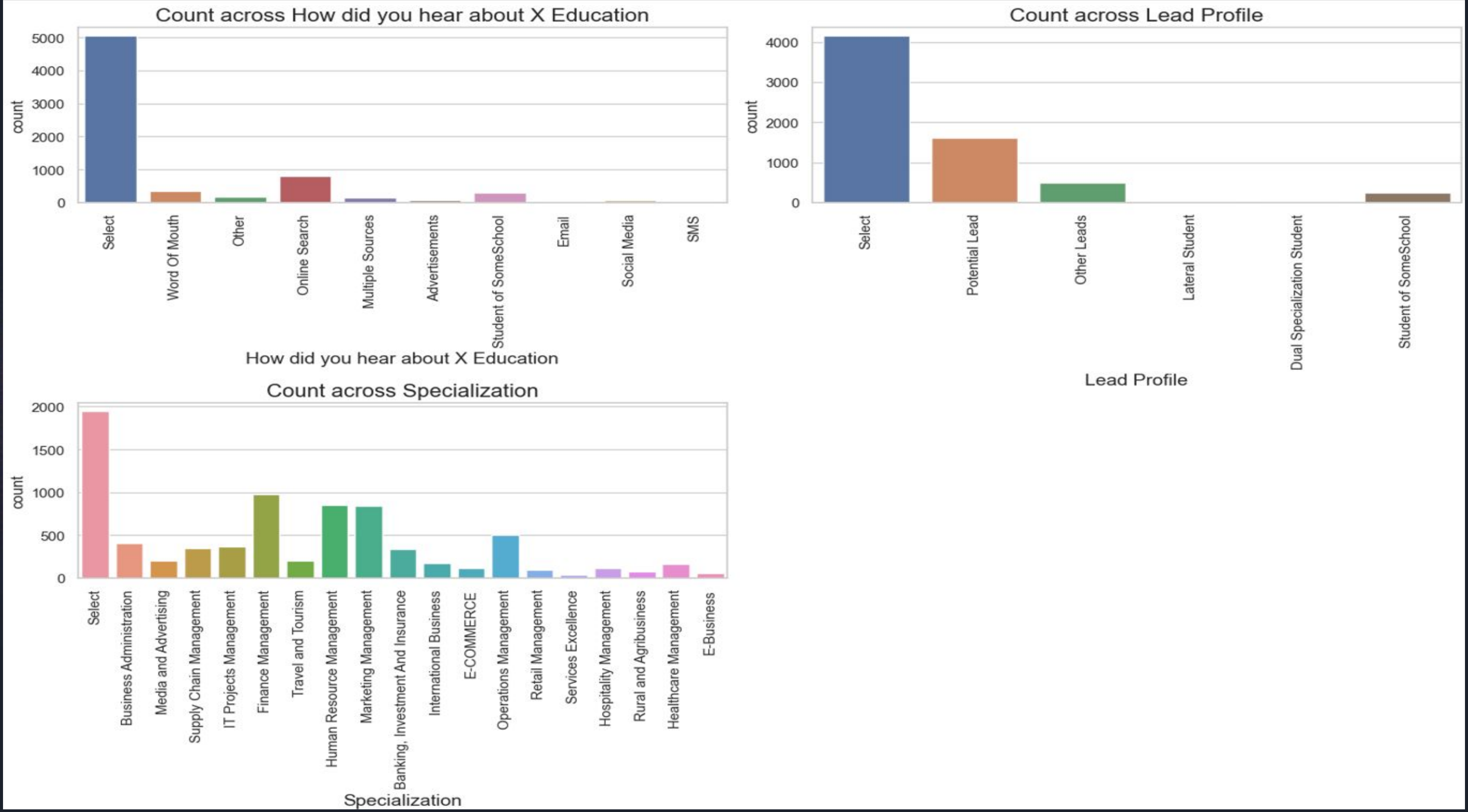
- Lead X wants a model that assigns each lead a score between 0-100 to identify hot leads and boost conversion rates.
- The CEO aims for a lead conversion rate of 80%.
- The model should be adaptable to future needs, such as managing peak times, optimizing manpower use, and strategies for after achieving targets.

Problem Approach

- Importing the data and inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE Rsquared VIF and p-values)
- Model Evaluation
- Making predictions on test set

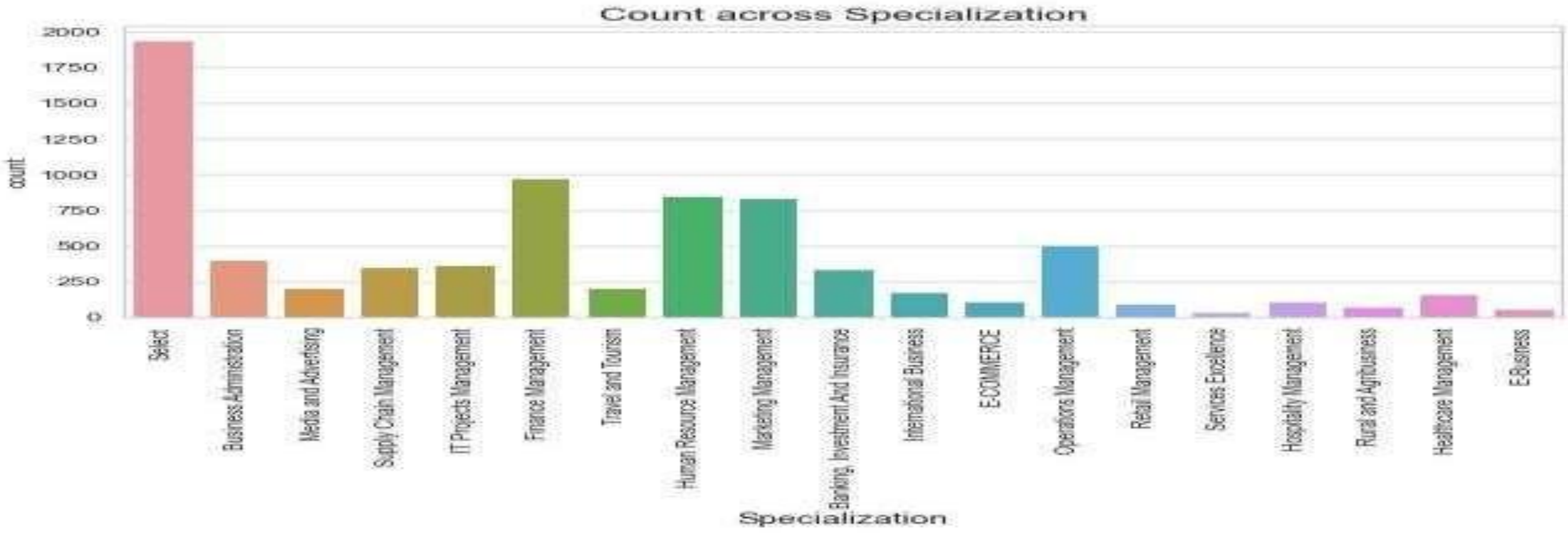
EDA – Data Cleaning

- There are a few columns in which there is a level called 'Select' which is taking care



Specialization

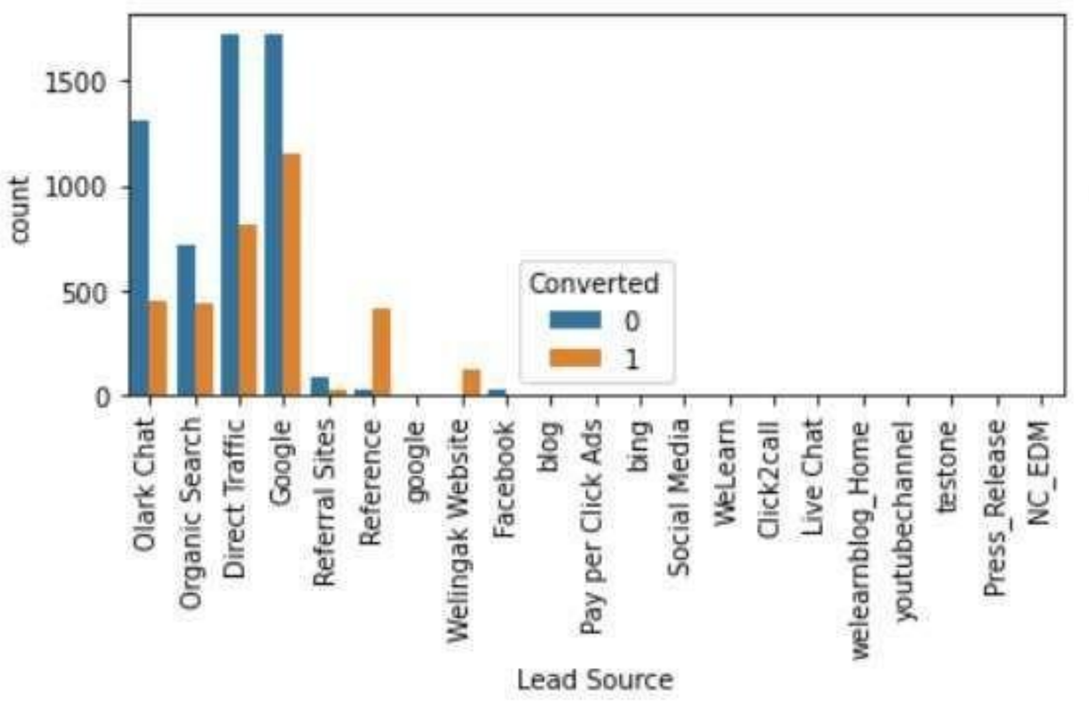
Leads from HR, Finance & Marketing management specializations are high probability to convert

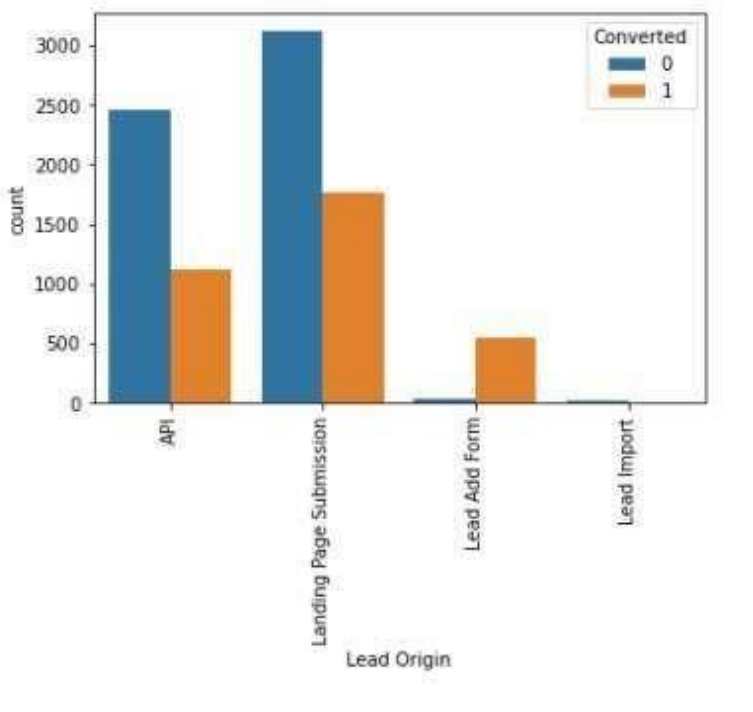


Lead Source & Lead origin

In lead source the leads through google & direct traffic high probability to convert

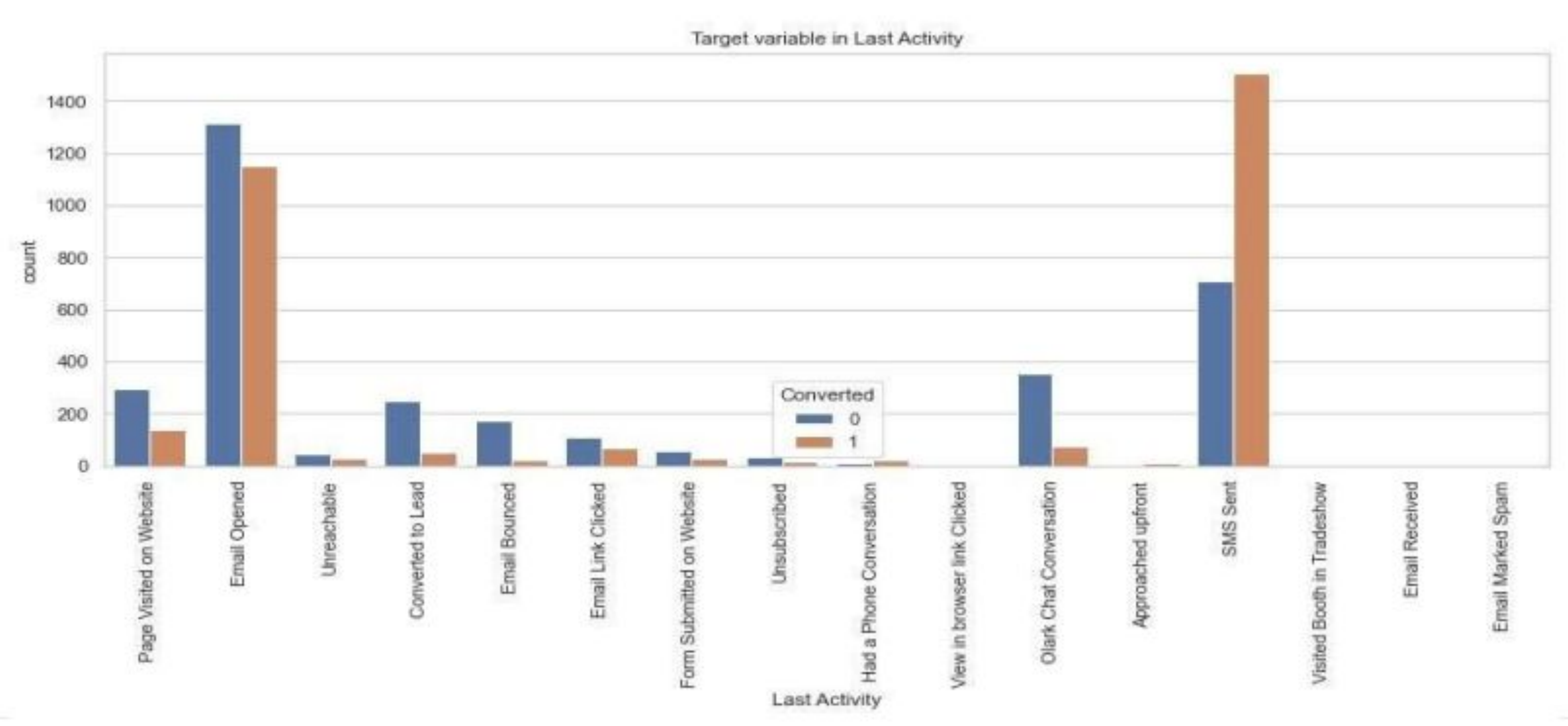
Whereas in Lead origin most number of leads are landing on submission





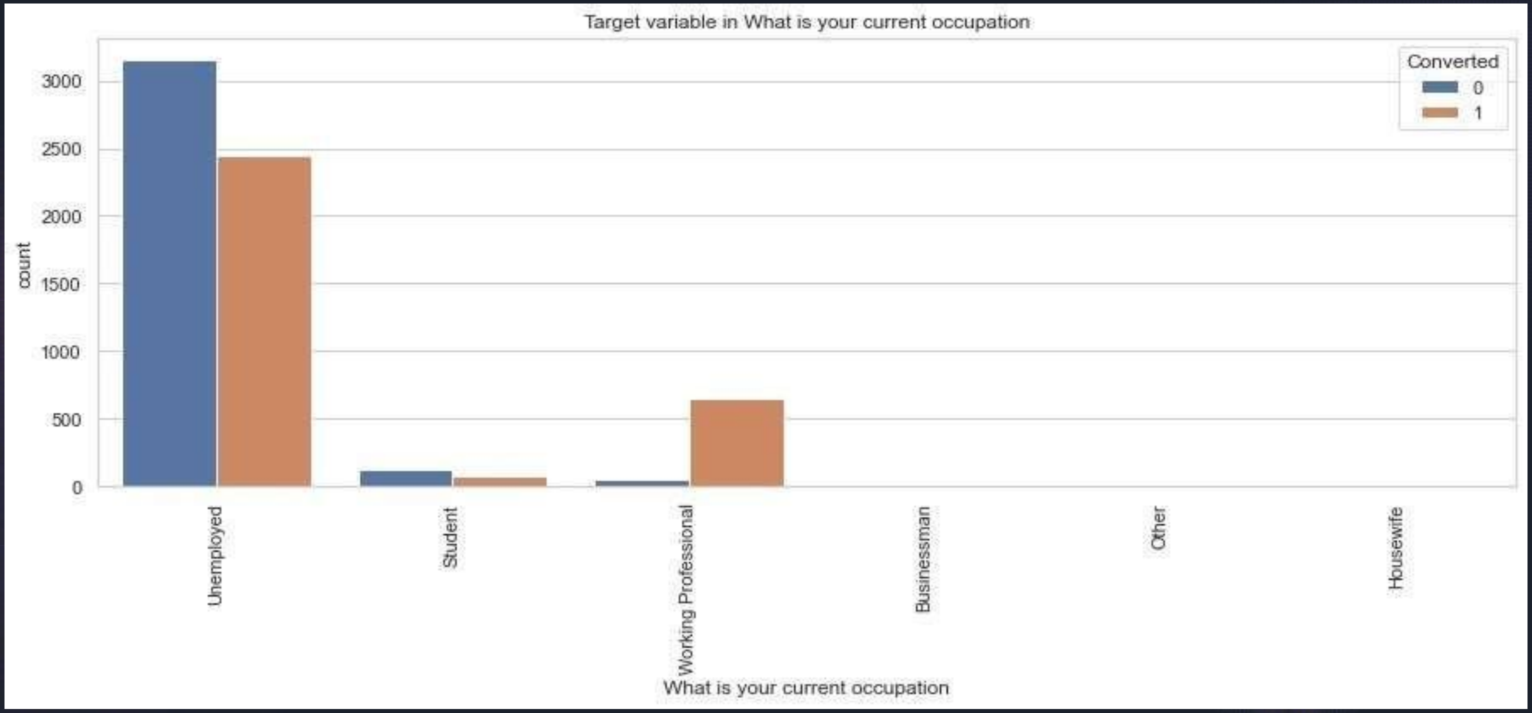
Last lead Activity

Leads which are opening email have high probability to convert, Same as Sending SMS will also benefit.



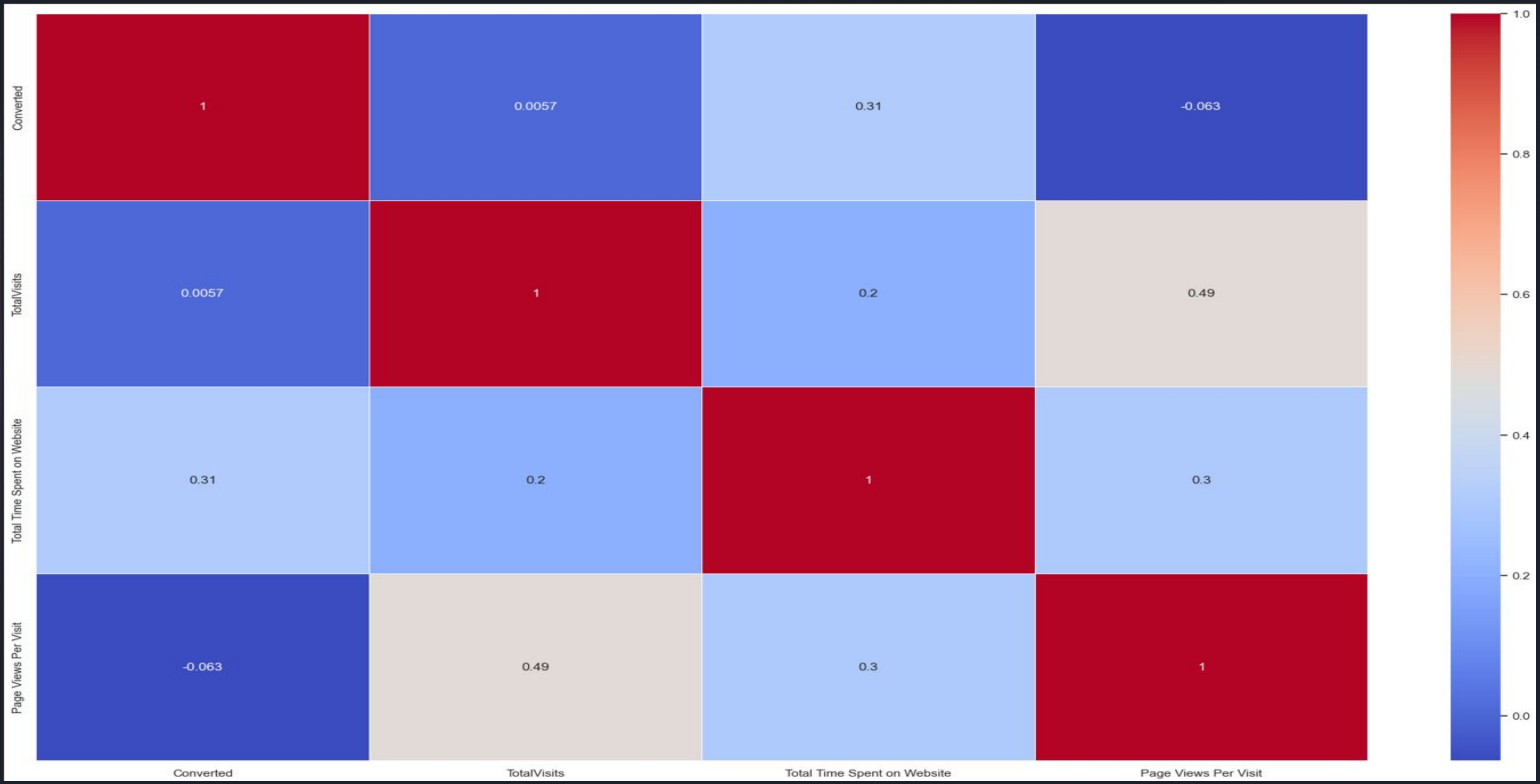
Last What is Your Occupation

Leads which are Unemployed are more interested to join the course than others.



Correlation

There is no correlation between the variables

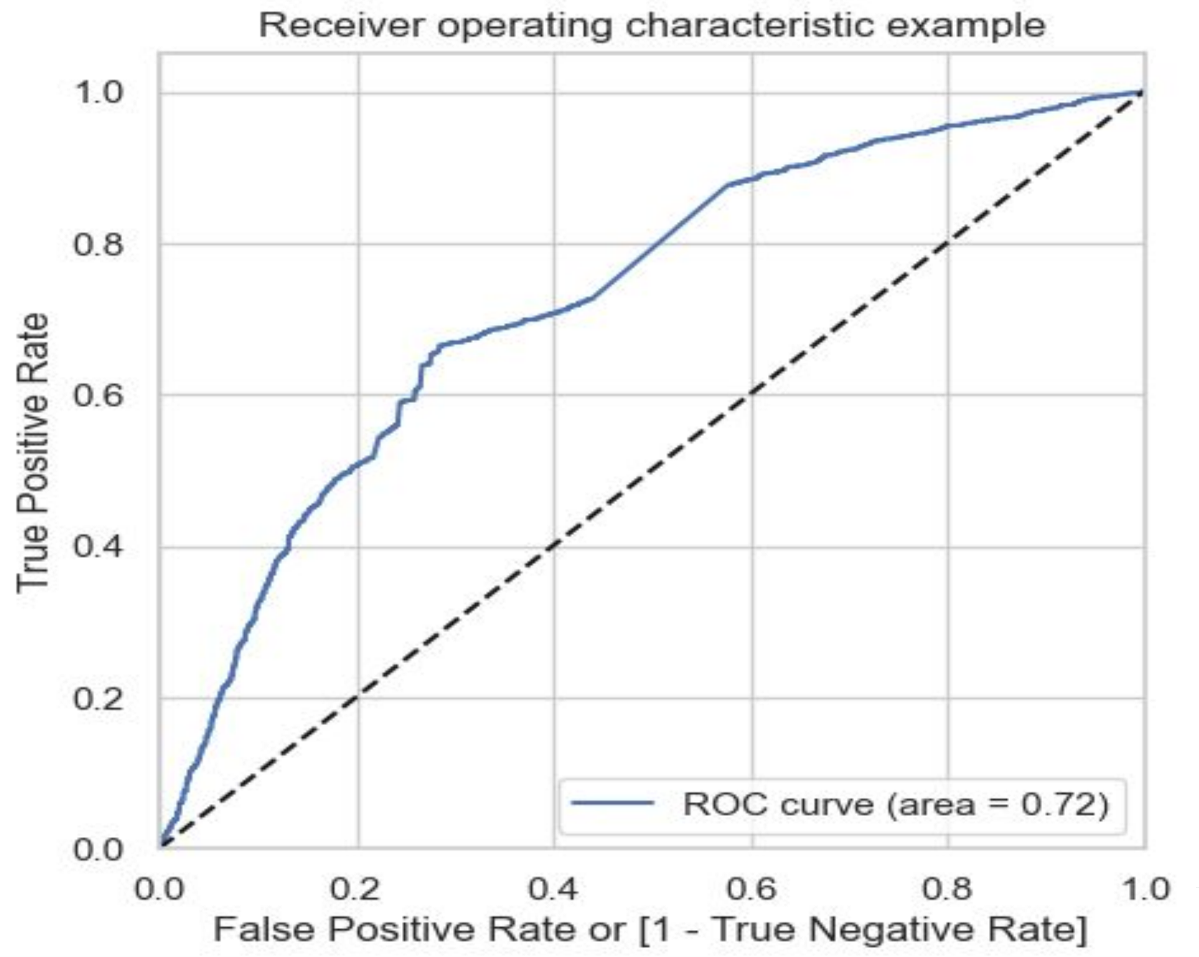


	Converted	TotalVisits	Total Time Spent on Website	Page Views Per Visit
Converted	1	0.0057	0.31	-0.063
TotalVisits	0.0057	1	0.2	0.49
Total Time Spent on Website	0.31	0.2	1	0.3
Page Views Per Visit	-0.063	0.49	0.3	1

Model Evaluation

ROC curve

0.42 is the tradeoff between Precision and Recall -
Thus we can safely choose to consider any Prospect Lead with Conversion **Probability higher than 42 % to be a hot Lead**



Observations

Train Data:

Accuracy : 66%
Sensitivity : 69%
Specificity : 63%

Test Data:

Accuracy : 67%
Sensitivity :51%
Specificity : 69%

Final Features list:

- Lead Source_Olark Chat
- Specialization_Others
- Lead Origin_Lead Add Form
- Lead Source_Welingak Website
- Total Time Spent on Website
- Lead Origin_Landing Page Submission
- What is your current occupation_Working Professionals
- Do Not Email

Conclusion

- We notice that the conversion rate is 30-35% (around average) for API and landing page submissions, but it's very low for Lead Add forms and Lead imports. This suggests we should focus more on leads coming from API and landing page submissions.
- Most leads come from Google and direct traffic. The highest conversion rates are achieved through referrals and the Welingak website.
- Leads who spend more time on the website are more likely to convert.
- The most common last activity is opening an email. The highest conversion rate comes from SMS messages. Most leads are unemployed, but the highest conversion rate is with working professionals.